International Journal of Applied Sciences, Engineering and Management
ISSN 2320 – 3439, Vol. 04, No. 01, January 2015, pp. 30 – 35

Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data

T. Latha[1], V P S Vinay Kumar[2]
[1]Student, Dept of CSE, Amrita Sai Institute of Science & Technology, Paritala, Krishna Dt
[2]Assistant Professor, Dept of CSE, Amrita Sai Institute of Science & Technology, Paritala, Krishna Dt
Email: latha.thota14@gmail.com, vinaykumar.csit@gmail.com

Abstract: Fine-grained multi-keyword search schemes over encrypted cloud data. Our innovative donations are three-fold. First, we commence the application scores and partiality factors upon keywords which facilitate the particular keyword search and modified user familiarity. we auxiliary take up the confidential sub-dictionaries procedure to accomplish better effectiveness on index structure, trapdoor generating and question. Lastly, we evaluate the sanctuary of the projected schemes in stipulations of discretion of credentials, privacy fortification of manifestation and trapdoor, and unlink capability of trapdoor. Through general experiments using the real-world dataset, we confirm the concert of the projected schemes. Both the safekeeping analysis and tentative results express that the projected schemes can accomplish the same protection level comparing to the presented ones and better routine in terms of functionality, query complication and competence.

Index Terms: *Searchable encryption, Multi-keyword, Fine-grained, Cloud computing.*

1. INTRODUCTION

Transmitting the information to the cloud servers. The data encryption, though, would considerably lower the usability of data outstanding to the complexity of penetrating over the encrypted data purely encrypting the statistics may still basis other sanctuary concerns. For example, Google Search uses SSL (Secure Sockets Layer) to encrypt the association among search user and Google server when confidential data, such as credentials and emails, appear in the search results. Nevertheless, if the explore user clicks into a different website as of the search consequences page, that website may be talented to categorize the explore terms that the user has worn. Firstly, the statistics owner needs to produce numerous keywords according to the outsourced data. These keywords are then encrypted and stored at the cloud server. When a explore user requirements to admission the outsourced data, it can select some appropriate keywords and send the nothing text of the preferred keywords to the cloud server. The cloud server then uses the cipher text to match the outsourced encrypted keywords, and lastly returns the matching results to the search user. To achieve the similar search efficiency and precision over encrypted data as that of plaintext keyword search, an extensive body of research has been developed in literature. propose a multi-keyword text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query. Yu et al. propose a multi-keyword top-k retrieval scheme which uses fully homomorphism encryption to encrypt the index/trapdoor and guarantees high security. Cao et al. propose a multi-keyword ranked search (MRSE), which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords. Although many search functionalities have been developed in previous literature towards precise and efficient searchable encryption, it is still difficult for searchable encryption to achieve the same user experience as that of the plaintext search, like Google search. The relevance scores of keywords can enable more precise returned results, and the preference factors of keywords represent the importance of keywords in the search keyword set specified by search users and correspondingly enables personalized search to cater to specific user preferences. It thus further improves the search functionalities and user experience.

2. SYSTEMMODEL, THREAT MODEL AND SECURITY REQUIREMENTS

2.1 System Model

We consider a system consists of three entities. *Data owner*: The data owner outsources her data tothe cloud for convenient and reliable data access to the corresponding search users. To protect the data privacy, the data owner encrypts the original data through symmetric encryption. To improve the search efficiency, the data owner generates some keywords for each outsourced document. The corresponding index is then created according to the keywords and a secret key. After that, the data owner sends the encrypted documents and the corresponding indexes to the cloud, and sends the symmetric key and secret key to search users. *Cloud server*: The cloud server is an intermediate entity which stores the encrypted documents and corresponding indexes that are received from the data owner, and provides data access and search services to search users. When a search user sends a keyword trapdoor to the cloud server, it would return a collection of matching documents based on certain operations.

• *Search user*: A search user queries the outsourced documents from the cloud server with following three steps. First, the search user receives both the

secret key and symmetric key from the data owner. Second, according to the search keywords, the search user uses the secret key to generate trapdoor and sends it to the cloud server. Last, she receives the matching document collection from the cloud server and decrypts them with the symmetric key.

2.2ThreatModel and Security Requirements

In our threat model, the cloud server is assumed to be "honest but- curious", which is the same as most related works on secure cloud data search. Specifically, the cloud server honestly follows the designated protocol specification. However, the cloud server could be "curious" to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information. We consider two threat models depending on the information available to the cloud server.

3. SYSTEM ARCHITECTURE

Fig-1. System Model

Known Ciphertext Model: The cloud server can only know encrypted document collection *C* and index collection, which are both outsourced from the data owner. Known Background Model: The cloud server can possess more knowledge than what can be accessed in the known ciphertext model, such as the correlation relationship of trapdoors and the related statistical of other information, i.e., the cloud server can possess the statistical information from a known comparable dataset which bears the similar nature to the targeting dataset. Similar to we assume search users are trusted entities, and they share the same symmetric key and secret key. Search users have pre-existing mutual trust with the data owner. For ease of illustration, we do not consider the secure distribution of the symmetric key and the secret key between the data owner and search users; it can be achieved through regular authentication and secure channel establishment protocols based on the prior security context shared between search users and the data owner. In addition, to make our presentations more focused, we do not consider following issues, including the access control problem on managing decryption capabilities given to users and the data collection's updating problem on inserting new documents, updating existing documents, and deleting existing documents, are separated issues. The interested readers on above issues may refer to. Based on the above threat model, we define the security requirements as follows:

• *Confidentiality of documents*: The outsourced documents provided by the data owner are stored in the cloud server. If they match the search keywords, they are sent to the search user. Due to the privacy of documents, they should not be identifiable except by the data owner and the authorized search users.

• *Privacy protection of index and trapdoor*: As discussed in Section 2.1, the index and the trapdoor are created based on the documents' keywords and the search keywords, respectively. If the cloud server identifies the content of index or trapdoor, and further deduces any association between keywords and encrypted documents, it may learn the major subject of a document, even the content of a short document. Therefore, the content of index and trapdoor cannot be identified by the cloud server.

• *Unlinkability of trapdoor*: The documents stored in the cloud server may be searched many times. The cloud server should not be able to learn any keyword information according to the trapdoors, e.g.,to determine two trapdoors which are originated from the same keywords. Otherwise, the cloud server can deduce relationship of trapdoors, and threaten to the privacy of keywords. Hence the trapdoor generation function should be randomized, rather than deterministic. Even in case that two search keyword sets are the same, the trapdoors should be different.

4. SYSTEM ANALYSIS

EXISTING SYSTEM: Using cloud computing, individuals can store their data on remote servers and allow data access to public users through the cloud servers. As the outsourced data are likely to contain sensitive privacy information, they are typically encrypted before uploaded to the cloud. This, however, significantly limits the usability of outsourced data due to the difficulty of searching over the encrypted data performance in terms of functionality, query complexity and efficiency. However, most existing proposals can only enable search with single logic operation, rather than the mixture of multiple logic operations on keywords, which motivates our work.

PROPOSED SYSTEM: this paper, we deal with this issue by increasing the fine-grained multi-keyword search schemes over encrypted cloud data. Our creative donations are three-fold. First, we commence the importance scores and predilection

factors upon keywords which facilitate the defined keyword search and adapted user familiarity. Second, we increase a matter-of-fact and very professional multi-keyword search scheme. The projected method can sustain difficult logic investigate the mixed "AND", "OR" and "NO" operations of keywords. Third, we additional utilize the confidential sub-dictionaries practice to accomplish better competence on guide construction, trapdoor generating and doubt. Lastly, we investigate the sanctuary of the wished-for schemes in provisos of discretion of credentials, privacy protection of index and trapdoor, and unlink capability of trapdoor. Through general experiments using the real-world dataset, we confirm the concert of the planned schemes. Both the defense analysis and tentative results display that the projected schemes can realize the same defense level comparing to the obtainable ones and better presentation in provisos of functionality, query complexity and efficiency.

Proposed system algorithms: The data owner firstly utilizes symmetric encryption algorithm. The security of this encryption algorithm has been proved in the known cipher text model. Thus, the content of index and trapdoor cannot be identified. Therefore, privacy protection of index and trapdoor can be achieved. Despite of the various advantages of cloud services, outsourcing sensitive information such as e-mails, personal health records, company finance data, government documents, etc.

RELATED WORK

This is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into account for developing the proposed system.

There are mainly two types of searchable encryption in literature, Searchable Public-key Encryption (SPE) and Searchable Symmetric Encryption (SSE).

SPE (Searchable Public-key Encryption)

SPE is first proposed by Boneh et al which supports single keyword search on encrypted data but the computation overhead is heavy. In the framework of SPE, Boneh et al. propose conjunctive, subset, and range queries on encrypted data. Hwang et al. propose a conjunctive keyword scheme which supports multi-keyword search. Zhang et al. propose an efficient public key encryption with conjunctive subset keywords search. However, these conjunctive keywords schemes can only return the results which match all the keywords simultaneously, and cannot rank the returned results. Qin et al. propose a ranked query scheme which uses a mask matrix to achieve cost-effectiveness. Yu et al. propose a multi-keyword top-k retrieval scheme with fully homomorphic encryption, which can return ranked results and achieve high security. In general, although SPE allows more expressive queries than SSE, it is less efficient, and therefore we adopt SPE in the work.

SSE (Searchable Symmetric Encryption)

The concept of SSE is first developed by Song et al. Wang et al. develop the ranked keyword search scheme, which considers the relevance score of a keyword. However, the above schemes cannot efficiently support multi-keyword search which is widely used to provide the better experience to the search user. Later, Sun et al. propose a multi keyword search scheme which considers the relevance scores of keywords, and it can achieve efficient query by utilizing the multidimensional tree technique. A widely adopted multi keyword search approach is multi-keyword ranked search (MRSE). This approach can return the ranked results of searching according to the number of matching keywords. Li et al. utilize the relevance score and k-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. Within this framework, they leverage an efficient index to further improve the search efficiency, and adopt the blind storage system to conceal access pattern of the search user. Li et al. also propose an authorized and ranked multi keyword search scheme (ARMS) over encrypted cloud data by leveraging the cipher text policy attribute-based encryption (CP-ABE) and SSE techniques. Security analysis demonstrates that the proposed ARMS scheme can achieve collusion resistance. In this paper, we propose FMS(CS) schemes which not only support multi-keyword search over encrypted data, but also achieve the fine-grained keyword search with the function to investigate the relevance scores and the preference factors of keywords and, more importantly, the logical rule of keywords. In addition, with the classified sub-dictionaries, our proposal is efficient in terms of index building, trapdoor generating and query.

5 .PROBLEM DEFINITION

Notations and Preliminaries

To solve this problem, it extends the index and inserts a random number "j which follows a normal distribution and can confuse the values of P •Q. Thus, enhanced MRSE can resist scale analysis attack. However, the introduction of "j causes precision decrease of the returned results. There is a trade-off between precision and security in MRSE. In comparison, our schemes do not suffer the scale analysis attack. Because the values of P • Q in our

schemes do not disclose any information due to the randomly selected sequences mentioned in Section and Section Therefore, our proposal can achieve the security without sacrificing precision. we do not consider following issues, including the access control problem on managing decryption capabilities given to users and the data collection's updating problem on inserting new documents, updating existing documents, and deleting existing documents, are separated issues.

6. MODULE DESCRIPTION

6.1 Searchable encryption:

This modules used on search the key word for encrypted text. This Module Used On Secure Purpose. More Secure For Encrypted Encryption to achieve the same user experience as that of the plaintext search, like Google search. This mainly attributes to following two issues. Firstly, query with user preferences is very popular in the Chipper text search

6.2 Multi-keyword:

Text search scheme which considers the relevance scores of keywords and utilizes a multidimensional tree technique to achieve efficient search query. Multi keyword top-k retrieval scheme which uses fully homomorphic encryption to encrypt the index/trapdoor and guarantees high security. Cao et al. [6] propose a multi-keyword ranked search (MRSE), which applies coordinate machine as the keyword matching rule, i.e., return data with the most matching keywords Fine-grained Multi-keyword Search (FMS) schemes over encrypted cloud data. multi-keyword search and coordinate matching using secure kNN computation scheme multi-keyword top-k retrieval scheme with fully homomorphic encryption, which can return ranked results and achieve high security.

6.3 Fine-grained:

we propose FMS(CS) schemes which not only support multi-keyword search over encrypted data, but also achieve the fine-grained keyword search with the function to investigate the relevance scores and the preference factors of keywords and, more importantly the logical rule of keywords. In addition, with the classified sub-dictionaries, our proposal is efficient in terms of index building, trapdoor generating and query. Fine-grained operations of keyword search, i.e.,"AND", "OR" and "NO" operations in Google Search, which are definitely practical and significantly enhance the functionalities of encrypted keyword search.

6.4 Cloud Computing:

Cloud computing is a computing term or metaphor that evolved in the late 1900s, based on utility and consumption of computer resources. Cloud computing involves deploying groups of remote servers and software networks that allow different kinds of data sources be uploaded for real time processing to generate computing results without the need to store processed data on the cloud. Clouds can be classified as public, private or hybrid. *Synonym expansion* are words with the same or similar meanings. In order to improve the accuracy of search results, the A Secure and Dynamic Multi-keyword Ranked extracted from out sourced text documents need to be extended by common synonyms, as cloud customers' searching input might be the synonyms of the predefined A Secure and Dynamic Multi-keyword Ranked, not the exact or fuzzy matching A Secure and Dynamic Multi-keyword Ranked due to the possible synonym substitution and/or her lack of exact knowledge about the data.

A common synonym thesaurus is built on the foundation of the New American Roget's College Thesaurus (NARCT) [14]. Then the keyword set is extended by using the constructed synonym thesaurus. Cryptography The art of protecting information by transforming it (*encrypting* it) into an unreadable format, called cipher text. Only those who possess a secret key can decipher (or *decrypt*) the message into plain text. encrypted messages can sometimes be broken by cryptanalysis, also called code *breaking*, although modern cryptography techniques are virtually unbreakable.

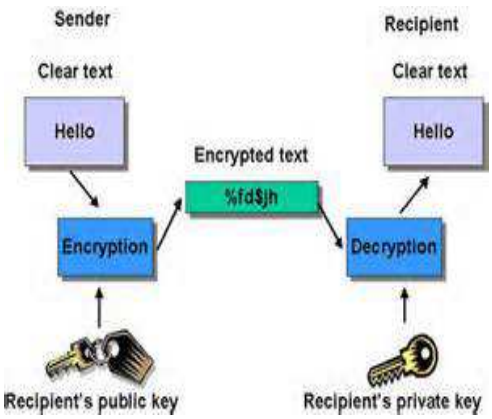

Fig-2: Encryption and Decryption

Encryption: In an encryption scheme, the message or information (referred to as *plaintext*) is encrypted using an encryption algorithm, turning it into an unreadable *cipher text* (ibid.). This is usually done with the use of an *encryption key*, which specifies how the message is to be encoded. Any adversary that can see the cipher text, should not be able to determine anything about the original message.

Decryption*:* An authorized party, however, is able to decode the cipher text using a decryption algorithm, that usually requires a *secret decryption key*, That adversaries do not have access to. For

technical reasons, an encryption scheme usually needs a key-generation algorithm, to randomly produce keys . Hence, it is an especially important thing to explore an effective multi-keyword ranked searching service over encrypted outsourced data.

7. INPUT DESIGN AND OUTPUT DESIGN

7.1INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.

Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

7.1.2 OBJECTIVES:

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

7.2OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user.

Efficient and intelligent output design improves the system's relationship to help user decision-making.

Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements. Select methods for presenting information. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives. Convey information about past activities, current status or projections of the Future. Signal important events, opportunities, problems, or warnings. Trigger an action. Confirm an action.

8. CONCLUSION

We have examined on the fine-grained multi keyword search (FMS) subject over encrypted cloud data, and future two FMS schemes. The FMS I includes both the significance scores and the partiality factors of keywords to augment more accurate search and enhanced users' experience, in that order. The FMS II realize secure and competent search with realistic functionality, i.e., "AND", "OR" and "NO" operations of keywords. In addition, we have planned the better schemes behind confidential sub-dictionaries (FMSCS) to advance competence. For the future work, we propose to add expand the application to reflect on the extensibility of the file set and the multi-user cloud environments. Towards this trend, we have made some beginning consequences on the extensibility and the multiuser cloud environments. Another remarkable topic is to increase the greatly scalable searchable encryption to enable able explore on large realistic databases.

9. REFERENCE

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdpbased service model for interdomain resource allocation in mobile cloud networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 5, pp. 2222–2232, 2012.

[2] M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," *IEEE*

Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting geo distributed clouds for e-health monitoring system with minimum service delay and privacy preservation," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 430–439, 2014.

[4] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy preserving data aggregation without secure channel: multivariate polynomial evaluation," in *Proceedings of INFOCOM*. IEEE, 2013, pp.2634–2642.

[5] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in *Proceedings of GLOBCOM*. IEEE, 2014, to appear.

[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 222–233, 2014.

[7] https://support.google.com/websearch/answer/173733?hl=en.

[8] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceedings of S&P*. IEEE, 2000, pp. 44–55.

[9] R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multi keyword ranked query over encrypted data in cloud computing," *Future Generation Computer Systems*, vol. 30, pp. 179–190, 2014.

[10] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, "Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage," *IEEE Transactions on Emerging Topics in Computing*, 2014, DOI10.1109/TETC.2014.2371239.

[11] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in *Proceedings of ICDCS*. IEEE, 2010, pp. 253–262.

[12] A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill, "Order-preserving symmetric encryption," in *Advances in Cryptology-EUROCRYPT*.Springer, 2009, pp. 224–241.

[13] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," *IEEE Transactions on Parallel and Distributed Systems*, vol. DOI: 10.1109/TPDS.2013.282, 2013.

[14] J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li, "Towards secure multikeyword top-k retrieval over encrypted cloud data," *IEEE Transactions on Dependable and Secure Computing*, vol. 10, no. 4, pp. 239–250, 2013.

[15] A. Arvanitis and G. Koutrika, "Towards preference-aware relational databases," in *International Conference on Data Engineering (ICDE)*. IEEE, 2012, pp. 426–437.

About the Authors:

Ms. T. Latha Is a Student of Amrita Sai Institute of Science And Technology, Paritala, Krishna Dt, Andhra Pradesh, presently she is pursuing M.Tech (C.S.E) from AMRITA SAI college and her areas of interest are Cloud Computing and Network Security.

Mr. V P S Vinay Kumar is working as an Assistant Professor in Department of Computer Science and Engineering of Amrita Sai Institute of Science And Technology. Having 5 years of teaching experience and areas of interests are Operating System, Data Mining and Computer Networks